Classification of Bread Wheat Varieties and Their Yield Characters with the Common Vector Approach

M. Bilginer Gülmezoğlu, and Nurdilek Gülmezoğlu

***Abstract*—** In this study, first of all, six varieties (Demir-2000, Gün-91, İkizce-96, Mızrak, Seval, Tosunbey) of bread wheat are classified by using the common vector approach (CVA). For this purpose, five yield characters (length of spike, spikelet number per spike, grain number per spike, spike weight, plant height) of varieties are used. These characters include 20 plant samples which were obtained from field studies conducted during two years under dry farming. Secondly, five yield characters taken from six varieties are classified by using CVA. Satisfactory results were obtained especially in the training set.

***Keywords*—** Character classification, common vector approach, wheat classification.

I. Introduction

CLASSIFICATION of many different plants or varieties of any plant is very important in many applications. Therefore, classification studies of varieties belonging to any plant have been conducted since five decades. Wheat varieties were classified using morphological methods at the beginning of these studies. Symes [1] classified Australian wheat varieties based on the granularity of their whole meal. Protein content and kernel hardness of wheat varieties were also used as quality evaluations in the classification of wheat varieties [2].

Computer-based algorithms have been extensively used in agriculture in order to classify various plants and their characters or samples. Classification of wheat varieties with computer algorithms has been become popular in recent years. Different methods were used in order to derive features or parameters from wheat varieties [3]–[8]. Some of these parameters are kernel color, hardness, image spots, dust particles, electrophorogram, plan-form spatial shape features and Fourier descriptors of kernel perimeters. Many classification algorithms such as Principal Component Analysis (PCA) [9]–[11], Linear Discriminant Analysis (LDA) [10]–[14], Neural Networks (NN) [15]–[17], Artificial Neural Network [15]–[17], and cluster analysis [8], [10], [19] have been used in classification of wheat varieties.

Delwiche and Massie classified two wheat classes by using partial least squares and multiple linear regression analyses in order to develop binary decision model for various combinations of two wheat classes [6]. Instead of samples taken from wheats, image analysis was also used in the wheat classification with crush-force parameters [4] and with statistical filters [5]. Neuman and Bushuk [7] applied digital image analysis for objective classification of wheat cultivars according to kernel type and identity.

In this study, we considered six varieties which are Demir, Gün-91, İkizce-96, Mızrak, Seval, Tosunbey of bread wheat. Each variety was represented with five yield characters which are length of spike, spikelet number per spike, grain number per spike, spike weight, plant height. Each character includes 20 plant samples which were taken from field studies conducted during two years under dry farming. There are two main purposes of our study: one is the variety classification and the other is the character classification. That is, the varieties and characters were applied to classification process separately. The common vector approach (CVA) was used in classification process. CVA is a well-known subspace method which was used in the classification of speech, speaker, human faces and motor faults [20]–[24]. Classification rates of varieties and characters are given in tables for training and testing stages.

II. Materials and Methods

Six bread wheat cultivars (Demir-2000, Gün-91, İkizce-96, Mızrak, Seval, Tosunbey are registered in Turkey) were planted in Eskisehir Osmangazi University research field, Turkey. The field experiment was conducted during growing season (2004/2005), seeded in October and harvested in July. The experiment was set up as completely randomized block design with four replications and a plot size of six rows 10 m long and a distance between rows of 25 cm. The soils at site of the experiments and precipitation means were sandy-loam and 288 mm, low in organic matter, and moderate for $CaCO_3$. At planting time, phosphorous and nitrogen were applied at a standard rate of 60 kg ha^{-1} P_2O_5 and 40 kg ha^{-1} N. Nitrogen topdressing treatments were applied at the tillering stage in March to a total of 70 kg N ha^{-1}. The plant height, length of spike, spikelet number per spike, number of grain per spike

M. Bilginer Gülmezoğlu is with Department of Electrical and Electronics Engineering, Faculty of Engineering,and Architecture, Eskisehir Osmangazi University, 26480, Eskisehir, Turkey (phone: +902222393750; e-mail: bgulmez@ogu.edu.tr).

Nurdilek Gülmezoğlu is with Department Soil Science and Plant Nutrition, Faculty of Agriculture, Eskisehir Osmangazi University, 26480, Eskisehir, Turkey (e-mail: dgulmez@ogu.edu.tr).

and spike weight were determined on randomly selected plants from each plot.

As in all classification methods, CVA has training and testing stages. In the training stage, a common vector which represents common or invariant properties of each class is calculated and an in difference subspace for each class is constructed. Let the vectors $a_1^c, a_2^c, ..., a_m^c \in R^n$ be the feature vectors for a variety-class C in the training set where $m \leq n$. Then each of these feature vectors which are assumed to be linearly independent can be written as

$$\mathbf{a}_i^c = \mathbf{a}_{i,dif}^c + \mathbf{a}_{com}^c + \varepsilon_i^c \quad \text{for } i=1,2,\ldots,m \tag{1}$$

where the vector $\mathbf{a}_{i,dif}^c$ indicates the differences resulting from climatic effects and alien-pollination, and the vector $\mathbf{a}_{com}^c$ is the common vector of the variety or character class C, and ε_i^c represents the error vector [22]. The common vector can be obtained from the subspace method. Let us define the covariance matrix of the feature vectors belonging to a variety or character class as

$$\mathbf{\Phi} = \sum_{i=1}^{m} (\mathbf{a}_i^c - \mathbf{a}_{ave}^c)(\mathbf{a}_i^c - \mathbf{a}_{ave}^c)^T \tag{2}$$

where $\mathbf{a}_{ave}^c$ is the average feature vector of Cth class whose covariance matrix is to be calculated and T indicates the transpose of a matrix.

The eigenvalues of the covariance matrix $\mathbf{\Phi}$ are non-negative and they can be written in decreasing order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Let $u_1^c, u_2^c, ..., u_n^c$ be the orthonormal eigenvectors corresponding to these eigenvalues. The first (m-1) eigenvectors of the covariance matrix corresponding to the nonzero eigenvalues form an orthonormal basis for the difference subspace B [22]. The orthogonal complement, $B^{\perp}$, is spanned by all the eigenvectors corresponding to the zero eigenvalues. This subspace is called the indifference subspace and has a dimension of (n-m+1). The direct sum of two subspaces B and $B^{\perp}$ is the whole space, and the intersection of them is the null space. The common vector can be shown as the linear combination of the eigenvectors corresponding to the zero eigenvalues of $\mathbf{\Phi}$ [22], that is,

$$\mathbf{a}_{com}^c = \left\langle \mathbf{a}_i^c, \mathbf{u}_m^c \right\rangle \mathbf{u}_m^c + \cdots + \left\langle \mathbf{a}_i^c, \mathbf{u}_n^c \right\rangle \mathbf{u}_n^c \ \forall\ i=1,2,\ldots,m \tag{3}$$

From here, the common vector acom is the projection of any feature vector onto the indifference subspace $B^{\perp}$. The common vector represents the common properties or invariant features of the variety or character class C. The common vector is independent from index i. Therefore, the common vector is unique for each class and all the error vectors ε_i^c would be zero.

During the classification stage, the following decision criterion is used:

$$distance = \underset{1 \leq C \leq S}{argmin} \left\| \sum_{j=m}^{n} \left\{ \left[\left(\mathbf{a}_x - \mathbf{a}_{ave}^c \right)^T \mathbf{u}_j^c \right] \mathbf{u}_j^c \right\} \right\|^2 \tag{4}$$

where $\mathbf{a_x}$ is an unknown or test vector and S indicates the total number of classes. If the distance is minimum for any class C, the feature vector $\mathbf{a_x}$ is assigned to class C.

III. Results

In the first study, each of six varieties forms one class in the CVA method. Five characters each of which includes 20 samples for each wheat variety form five feature vectors of that variety. Therefore, there are six classes and each class has five feature vectors. When the feature vectors (characters) used in the training stage were tested, all classes (varieties) were correctly classified, i.e., 100% correct recognition rate was obtained. When the "leave-one-out" strategy was used in the testing stage, that is, when four feature characters were used in the training stage and remaining one character was tested, 36.7% correct recognition rate was obtained as average of "leave-one-out" steps. The results obtained from this study are given in Table 1. The average score obtained in the test set is very low because samples included in the characters representing same variety are much different.

Table I
Correct Recognition Rates Of Wheat Varieties As Percentage

VARIETIES	TRAINING SET	TEST SET
Demir-2000	100	60
Gün-91	100	40
İkizce-96	100	40
Mızrak	100	40
Seval	100	20
Tosunbey	100	20
Average	100	36.7

In the second study, the characters were classified by using CVA. First of all, five characters were considered and each of five characters forms one class in the CVA method. Twenty samples taken from each variety for any character form one feature vector of that character. Therefore, there are five classes and each class has six feature vectors. When the feature vectors, each of them includes 20 samples, used in the training stage were tested, all classes (characters) were correctly classified, i.e., 100% correct recognition rate was obtained. When the "leave-one-out" strategy was used in the testing stage, that is, when five feature characters were used in the training stage and remaining one character was tested, 80% correct recognition rate was obtained as average of "leave-one-out" steps. The results obtained for this study are given in Table 2.

TABLE II
CORRECT RECOGNITION RATES OF FIVE YIELD CHARACTERS AS PERCENTAGE

YIELD CHARACTERS	TRAINING SET	TEST SET
Length of spike	100	33.3
Spikelet number per spike	100	66.7
Number of grain per spike	100	100
Spike weight	100	100
Plant height	100	100
Average	100	80

Secondly, last four characters (spikelet number per spike, grain number per spike, spike weight, plant height) were classified. All characters were correctly classified (100% correct recognition rate was obtained) in both the training and testing stages. These scores are remarkable because samples taken from varieties for each character are close to each other and well represent that character. These results are given in Table 3.

TABLE III
CORRECT RECOGNITION RATES OF FOUR CHARACTERS AS PERCENTAGE

YIELD CHARACTERS	TRAINING SET	TEST SET
Spikelet number per spike	100	100
Number of grain per	100	100
Spike weight	100	100
Plant height	100	100
Average	100	100

IV. DISCUSSION

It is known that varieties of different plants have been successfully classified by using various computer-based algorithms. Classification of wheat varieties with computer algorithms has been become popular in recent years [4], [5], [7]. Therefore in this study, first of all, six wheat varieties were classified by using CVA method. In spite of 100% correct recognition rate in the training set, very low recognition rate (36.7%) was obtained in the test set. The reason is that samples included in the characters representing same variety are much different. Thus, common properties or invariant features of each variety can not be extracted and indifference subspace can not be constructed efficiently.

Additionally, characters were classified by using CVA method. Initially, five characters are applied to the classification process and 100% and 80% recognition rates were obtained for the training and test set respectively. The reason of low score for the test set is that the samples of length of spike and spikelet number per spike characters are very close to each other. Therefore these two characters are confused. When the length of spike character is discarded, that is, when last four characters are classified, all characters are correctly classified (100% recognition rate is obtained) in both the training and test sets.

V. CONCLUSION

It is concluded that the CVA method was very successful in the classification of different varieties belonging to any plant and/or of different characters belonging to any variety. Such classifications can be very helpful in assignment of unknown varieties or unknown characters to correct plant. As a future work, number of varieties for any plant and the number of characters will be increased. Satisfactory results are also expected from this work.

REFERENCES

[1] K J. Symes, Classification of Australian wheat varieties based on the granularity of their wholemeal. Australian Journal of Experimental Agriculture and Animal Husbandry 1(1) pp. 18-23, 1961. http://dx.doi.org/10.1071/EA9610018

[2] D. C. .Slaughter, K. H. Norris and W. R. Hruschka, Quality and classification of hard red wheat. Cereal Chemistry 69(4) pp. 428-432, 1992.

[3] D. Wang, F. E. Dowell and R. E. Lacey, Single wheat kernel color classification using neural networks. Transactions of the ASABE 42(1) pp. 233-240, 1999. http://dx.doi.org/10.13031/2013.13200

[4] I. Y. Zayas, C. R. Martin, J. L Steele and A. Katsevich, Wheat classification using image analysis and crush-force parameters. Transactions of the ASABE 39(6) pp. 2199-2204, 1996. http://dx.doi.org/10.13031/2013.27725

[5] H. Utku, and H. Köksel, Use of statistical filters in the classification of wheats by image analysis. Journal of Food Engineering 36(4) pp. 385-394, 1998. http://dx.doi.org/10.1016/S0260-8774(98)00072-7

[6] S. R. Delwiche, and D. R. Massie, Classification of wheat by visible and near-infrared reflectance from single kernels. Analytical Techniques and Instrumentation 73(3) pp. 399-405, 1996.

[7] M. Neuman and W. Bushuk, Discrimination of wheat class and variety by digital image analysis of whole grain samples. Journal of Cereal Science 6(2) pp. 125-132, 1987. http://dx.doi.org/10.1016/S0733-5210(87)80049-8

[8] M. Shuaib, M. Jamal, H. Akbar, I. Khan, and R. Khalid, Evaluation of wheat by polyacrylamide gel electrophoresis. African Journal of Biotechnology 9(2) pp. 243-247, 2010.

[9] J. J. Ordaz-Ortiz, M. F. Devaux and L. Saulnier, Classification of wheat varieties based on structural features of arabinoxylans as revealed by endoxylanase treatment of flour and grain. Journal of Agriculture and Food Chemistry 53(21) pp. 8349-8356, 2005. http://dx.doi.org/10.1021/jf050755v

[10] P. Zapotoczny, M. Zielinska and Z. Nita, Application of image analysis for the varietal classification of barley: Morphological features Journal of Cereal Science 48(1) pp. 104-110, 2008. http://dx.doi.org/10.1016/j.jcs.2007.08.006

[11] F. Kraic, J. Mocak, T. Rohacik and J. Sokolovicova, Chemometric characterization and classification of new wheat genotypes. Nova Biotechnologica 9(1) pp. 101-106, 2009.

[12] C. Armanino, R. D. Acutis and M. R. Festa, Wheat lipids to discriminate species, varieties, geographical origins and crop years. Analytica Chimica Acta 454(2) pp. 315-326, 2002. http://dx.doi.org/10.1016/S0003-2670(01)01548-3

[13] S. R. Delwiche, High-speed bichromatic inspection of wheat kernels for mold and color class using high-power pulsed LEDs. Sensing and Instrumentation for Food Quality and Safety 2(2) pp. 103-110, 2008. http://dx.doi.org/10.1007/s11694-008-9037-1

[14] Z. Sramkova, F. Kraic, J. Jurovata, E. Gregova and E. Sturdik, Chemometric analysis of nutriti onal bread-making quality attributes of wheat cultivars. Acta Chimica Slovaca 2(2) pp. 139-146, 2009.

[15] Y. R. Chen, R. Dewiche, and W. R. Hruschka, Classification of hard red wheat by feedforward backpropagation neural networks. Cereal Chemistry 72(3) pp. 317-319, 1995.

[16] H. A. Sorensen, M. M. Sperotto, M. Peterson, C. Keçmir, L. Radzikowski, S. Jacobsen and I. Sondergaard, Variety identification of wheat using mass spectrometry with neural networks and the influence of mass spectra processing prior to neural network analysis. Rapid Communications in Mass Spectrometry 16(12) pp. 1232-1237, 2002. http://dx.doi.org/10.1002/rcm.709

[17] I. Sondergaard, K. Jensen, and B. N. Krath, Classification of wheat varieties by isoelectric focusing patterns of gliadins and neural network. Electrophoresis 15(5) pp. 584-588, 1994. http://dx.doi.org/10.1002/elps.1150150181

[18] S. Jacobsen, L. Nesic, M. Petersen and I. Sondergaard, Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network. Electrophoresis 22(6) pp. 1242-1245, 2001. http://dx.doi.org/10.1002/1522-2683()22:6<1242::AID-ELPS1242>3.0.CO;2-Q

[19] N. A. R. Tahir, Assessment of genetic diversity among wheat varieties in sulaimanyah using random amplified polymorphic DNA (RAPD) analysis. Jordan Journal of Biological Sciences 1(4), pp. 159-194, 2008.

[20] M. B. Gülmezoğlu, and A. Barkana, Text-Dependent speaker recognition by Using Gram-Schmidt Orthogonalization Method. IASTED International Conference on Signal processing and Communications, Proc. 16. IASTED Int. Conference on Applied Mathematics, pp. 438-440, Canary Islands, Spain, 1998.

[21] M. B. Gülmezoğlu, V. Dzhafarov, M. Keskin. and A. Barkana, A Novel Approach to Isolated Word Recognition. IEEE Transactions on Speech and Audio Processing 7(6), pp. 620-628, 1999. http://dx.doi.org/10.1109/89.799687

[22] M. B. Gülmezoğlu, V. Dzhafarov and A. Barkana, The Common Vector approach and its Relation to Principal Component Analysis. IEEE Trans. Speech and Audio Processing 9(6), pp. 655-662, 2001. http://dx.doi.org/10.1109/89.943343

[23] M. B. Gülmezoğlu and S. Ergin, An Approach for Bearing Fault Detection in Electrical Motors. European Transactions on Electrical Power 17(6), pp. 628-641, 2007. http://dx.doi.org/10.1002/etep.161

[24] S. Ergin and M. B. Gülmezoğlu, Face Recognition Based on Face Partitions Using Common Vector Approach. Proc. of 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP'08), pp. 624-628, St. Julians, Malta, March, 2008. http://dx.doi.org/10.1109/ISCCSP.2008.4537300